Inference for GLM's

Assume $Y_i \overset{ind}{\sim} \text{Bernoulli}(\theta_i)$ $\forall i$ $\theta_i \in [0,1]$, the parameter space

where $\theta_i = f_{pr}(\vec{x}_i)$
which we approximate with $\phi(\vec{x}_i \vec{\beta})$
where $\phi : \mathbb{R} \to [0,1]$,
the link function

This should be called "Bernoulli Regression" but instead it's called by its link function

$$\vec{b} := \text{argmax}\left\{ \prod_{i=1}^{n} \phi(\vec{x}_i \cdot \vec{w})^{Y_i} (1-\phi(\vec{x}_i \cdot \vec{w}))^{1-Y_i} \right\} \Rightarrow \vec{b} \text{ is the MLE by definition}$$

MLE thm for vector estimators: if $\dim[\vec{\theta}] = k$

$$\sqrt{n}\left(\vec{\hat{\theta}}^{MLE} - \vec{\theta}\right) \xrightarrow{d} N_k\left(\vec{0}_k, \underbrace{I^{-1}(\vec{\theta})}_{\text{Inv. Fisher Information Matrix}}\right) \overset{\text{slutsky's}}{\Longrightarrow} \sqrt{n}\left(\vec{\hat{\theta}}^{MLE} - \vec{\theta}\right) \xrightarrow{d} N_k\left(\vec{0}_k, I^{-1}(\vec{\hat{\theta}}^{MLE})\right)$$

$$I(\vec{\theta}) := \text{Var}_{\vec{X}}\left[\frac{\partial}{\partial \vec{\theta}}\left[\ell(\vec{\theta}; \vec{X})\right]\right]$$

We won't cover this

$$\vec{B} \dot{\sim} N_{p+1}\left(\vec{\beta}, \frac{1}{n} I^{-1}(\vec{b})\right) \Rightarrow B_j \dot{\sim} N\left(\beta_j, \frac{1}{n I(\vec{b})_{j,j}^{-1}}\right)$$

The Fisher Information is computed by the computer based on link function

This gives us Wald tests for $H_0 : \beta_j = 0$

$$\frac{B_j}{\sqrt{\frac{I^{-1}(\vec{b})_{j,j}}{n}}} \dot{\sim} N(0,1)$$

How do we get the equivalent of F-tests? We can use the Wald χ^2. But usually people use LikR tests.

from 341

Recall the LikR test: $H_0: \vec{\theta} = \vec{\theta}_0$ where $\dim[\vec{\theta}] = k$ and $\vec{\theta}$ is unconstrained

$$\hat{LR} := \frac{\mathcal{L}(\hat{\vec{\theta}}^{mle}; \vec{x})}{\mathcal{L}(\vec{\theta}_0; \vec{x})} \geq 1$$ since the $\mathcal{L}(\hat{\vec{\theta}}^{mle}; \vec{x})$ is the max of the likelihood

$$\hat{\Lambda} := 2\ln(\hat{LR}) \,\dot\sim\, \chi^2_k$$

---

not from 341

Also, $H_0: \vec{\theta}_S = \vec{\theta}_{S_0}$ where $S \subseteq \{1, \ldots, k\}$ e.g. $S = \{1, 2\}$

and $A := \{0\} \cup S^c$

$$\hat{LR} = \frac{\mathcal{L}(\hat{\vec{\theta}}^{mle}; \vec{x})}{\mathcal{L}(\theta_{S_{0_1}}, \theta_{S_{0_2}}, \hat{\theta}_3^{mle}, \ldots, \hat{\theta}_k^{mle}; \vec{x})}$$

$$\hat{\Lambda} = 2\ln(\hat{LR}) \,\dot\sim\, \chi^2_{|S|}$$

How does that help us here? The omnibus test in logistic regression would be:

$$\hat{LR} = \frac{\prod_{i=1}^{n} \left(\frac{1}{1+e^{-\vec{x}_i \cdot \vec{b}}}\right)^{y_i} \left(\frac{1}{1+e^{\vec{x}_i \cdot \vec{b}}}\right)^{1-y_i}}{\prod_{i=1}^{n} \left(\frac{1}{1+e^{-\tilde{b}_0}}\right)^{y_i} \left(\frac{1}{1+e^{+\tilde{b}_0}}\right)^{1-y_i}}$$

then $2\ln(\hat{LR})$ is compared to $\chi^2_{p, 1-\alpha}$. If greater, reject H_0.

where $\tilde{b}_0 := \underset{w \in \mathbb{R}}{\text{argmax}} \prod_{i=1}^{n} \left(\frac{1}{1+e^{-w}}\right)^{y_i} \left(\frac{1}{1+e^{+w}}\right)^{1-y_i}$, the model where all β_j's $= 0$, thus we only fit the intercept

What about partial tests? $H_0: \vec{\beta}_S = \vec{0}_{|S|}$

$$\hat{LR} = \frac{\prod_{i=1}^{n} \left(\frac{1}{1+e^{-\vec{x}_i \vec{b}}}\right)^{y_i} \left(\frac{1}{1+e^{+\vec{x}_i \vec{b}}}\right)^{1-y_i}}{\prod_{i=1}^{n} \left(\frac{1}{1+e^{-\vec{x}_{A_i} \vec{b}_A}}\right)^{y_i} \left(\frac{1}{1+e^{+\vec{x}_{A_i} \vec{b}_A}}\right)^{1-y_i}}$$

then $2\ln(\hat{LR})$ is compared to $\chi^2_{S,1-\alpha}$. If greater $\Rightarrow$ Reject H_0

where $\vec{b}_A := \underset{\vec{w} \in \mathbb{R}^{|A|}}{\text{argmax}} \left\{ \prod_{i=1}^{n} \left(\frac{1}{1+e^{-\vec{x}_{A_i} \vec{w}}}\right)^{y_i} \left(\frac{1}{1+e^{+\vec{x}_{A_i} \vec{w}}}\right)^{1-y_i} \right\}$, the restricted model where all $\beta_j = 0$ for $j \in S$.

---

let $\mathcal{Y} = \{0, 1, 2, \ldots\}$. Models for the response are called "count models".

Which rv's have support $\mathcal{Y}$? Poisson, Negative Binomial are the most commonly used.

Recall, $E[Y_i] = \theta_i$, expectation here is range in $(0, \infty)$

Assume $Y_i \overset{ind}{\sim} \text{Poisson}(\theta_i)$, $\theta_i \in [0, \infty]$, the parameter space

where $\theta_i = f(\vec{x}_i)$ which we approximate with $\phi(\vec{x}_i \vec{\beta})$

where $\phi: \mathbb{R} \to [0, \infty)$, the link function. This is called "Poisson Regression".

Where Bernoulli Regression had many popular choices of link functions, here we use $\phi(u) = e^u$ most often. We won't consider others.

$$\Rightarrow \vec{b} = \underset{\vec{w} \in \mathbb{R}^{p+1}}{\text{argmax}} \left\{ \prod_{i=1}^{n} \frac{\phi(\vec{x}_i \vec{w})^{Y_i} e^{-\phi(\vec{x}_i \vec{w})}}{Y_i!} \right\} = \text{argmax} \left\{ \prod_{i=1}^{n} \phi(\vec{x}_i \vec{w})^{Y_i} e^{-\phi(\vec{x}_i \vec{w})} \right\}$$

the MLE for $\vec{\beta}$

$$= \text{argmax} \left\{ \sum Y_i \ln(\phi(\vec{x}_i \vec{w})) - \phi(\vec{x}_i \vec{w}) \right\} = \text{argmax} \left\{ \sum Y_i \vec{x}_i \vec{w} - e^{\vec{x}_i \vec{w}} \right\}$$

which has no closed form solution

Same as before

By MLE thm, and Slutsky's / CMT $\vec{B} \overset{\cdot}{\sim} N_{p+1}\left(\vec{\beta}, \frac{1}{n} I(\vec{\beta})^{-1}\right) \Rightarrow B_j \overset{\cdot}{\sim} N\left(\beta_j, \frac{1}{n I(\vec{\beta})^{-1}_{j,j}}\right)$

Fisher Information Matrix. Doable, but just not covered

Omnibus test, partial tests $\Rightarrow$ all done with LR test as before.

How does prediction work?

$e^{\vec{x}_* \vec{b}}$ returns the expected count for $\vec{x}_*$ which is $\in (0,\infty) \neq \{0,1,\dots\} = \mathcal{Y}$

If you need a prediction as a count, you can round so

$$\hat{y}_* = g(\vec{x}_*) = \text{round}\left(e^{\vec{x}_* \vec{b}}\right)$$

What does b_j represent?
a multiplicative factor change of e^{b_j} in the response

Not covered... Recall

$$Y \sim \text{ExtNegBin}(r, p) := \frac{\Gamma(y+r)}{y!\,\Gamma(r)} (1-p)^y p^r, \quad E(Y) = r\frac{1-p}{p}$$

Reparameterize to the mean

let $\theta = r\frac{1-p}{p} \Rightarrow \frac{\theta}{r} = \frac{1}{p} - 1 \Rightarrow \frac{\theta}{r} + 1 = \frac{1}{p} \Rightarrow \frac{\theta+r}{r} = \frac{1}{p} \Rightarrow p = \frac{r}{\theta+r} \Rightarrow 1-p = \frac{\theta}{\theta+r}$

$\Rightarrow Y \sim \text{ExtNegBin}(r, \theta) = \frac{\Gamma(y+r)}{y!\,\Gamma(r)} \left(\frac{\theta}{\theta+r}\right)^y \left(\frac{r}{\theta+r}\right)^r$, $E(Y) = \theta$, expectation lives in θ range is $(0,\infty)$

let $\theta_i = \phi(\vec{x}_i \vec{\beta}) = e^{\vec{x}_i \vec{\beta}}$

$$R, \vec{B} = \underset{\substack{r \in (0,\infty), \\ \vec{w} \in \mathbb{R}^{p+1}}}{\text{argmax}} \left\{ \prod_{i=1}^{n} \frac{\Gamma(Y_i + r)}{Y_i!\,\Gamma(r)} \left(\frac{\phi(\vec{x}_i \vec{w})}{\phi(\vec{x}_i \vec{w}) + r}\right)^{Y_i} \left(\frac{r}{\phi(\vec{x}_i \vec{w}) + r}\right)^r \right\} = \text{argmax} \left\{ \frac{\Gamma(Y_i + r)\, r^r\, \phi(\vec{x}_i \vec{w})^{Y_i}}{\Gamma(r) \left(\phi(\vec{x}_i \vec{w}) + r\right)^{Y_i + r}} \right\}$$

MLE's

r is a nuisance parameter

"Negative Binomial Regression"

no closed form

$\hat{y}_* = \text{round}\left(e^{\vec{x}_* \vec{b}}\right)$ just like before let $\mathcal{Y} = (0, \infty)$ survival/churn response

Consider $Y \sim \text{Weibull}(k, \lambda) = (k\lambda)(\lambda y)^{k-1} e^{-(\lambda y)^k} \mathbb{1}_{y>0}$, $E[Y] = \frac{1}{\lambda}\Gamma(1+\frac{1}{k})$

Let $\theta = \frac{1}{\lambda}$ reparameterize to get

$$Y \sim \text{Weibull}(k, \theta) = \frac{k}{\theta}\left(\frac{y}{\theta}\right)^{k-1} e^{-\left(\frac{y}{\theta}\right)^k} \mathbb{1}_{y>0}, \quad E(Y) = \theta\,\Gamma\left(1+\frac{1}{k}\right)$$

Notice expectation linear in θ and ranges in $(0, \infty)$

let $Y_i \overset{ind}{\sim} \text{Weibull}(k, \theta_i)$ where $\theta_i = \phi(\vec{x}_i \vec{\beta}) = e^{\vec{x}_i \vec{\beta}}$

$$\underbrace{k, \vec{b}}_{} = \underset{\vec{w} \in \mathbb{R}^{p+1}}{\text{argmax}} \left\{ \prod_{i=1}^{n} \frac{k}{e^{\vec{x}_i \vec{w}}} \left(\frac{Y_i}{e^{\vec{x}_i \vec{w}}}\right)^{k-1} e^{-\left(\frac{Y_i}{e^{\vec{x}_i \vec{w}}}\right)^k} \right\}$$

Weibull Regression

no closed form solution

MLE's,

k is nuisance!

Wald tests for single effects and χ^2 using likelihood ratio for multiple effects including omnibus tests.

What if there's censoring? No problem! let $\vec{c}$ be the dummy recording censoring

$$k, \vec{b} = \underset{\vec{w} \in \mathbb{R}^{p+1}}{\text{argmax}} \left\{ \prod_{\{i : c_i = 0\}} \frac{k}{e^{\vec{x}_i \vec{w}}} \left(\frac{Y_i}{e^{\vec{x}_i \vec{w}}}\right)^{k-1} e^{-\left(\frac{Y_i}{e^{\vec{x}_i \vec{w}}}\right)^k} \prod_{\{i : c_i = 1\}} e^{-\left(\frac{Y_i}{e^{\vec{x}_i \vec{w}}}\right)^k} \right\}$$

b_j represents multiplicative factor change in response

Prediction: $\hat{y}_* = g(\vec{x}_*) = e^{\vec{x}_* \vec{b}}\, \Gamma\left(1 + \frac{1}{\hat{k}_{MLE}}\right)$